

Chicago Conference of BAE
September 23-28, 1946
Chas. F. Sarle

For administrative use

SAMPLING IN AGRICULTURAL ECONOMIC RESEARCH AND STATISTICS

Introduction

This morning Dr. Elliott briefly described the proposed program of research and statistics of the Bureau of Agricultural Economics and suggested "2-way cooperation" with the Land Grant Colleges. In the field of Agricultural Estimates there is at the present time 2-way cooperation either with the State Departments of Agriculture and/or the Land Grant Colleges in 29 of the Field Offices representing 33 States. This cooperation is a natural outgrowth of the fact that the Federal Government is primarily interested in National Statistics with a breakdown by regions and States, whereas State Agencies are primarily interested in State Statistics with a breakdown by areas and counties within the State. The Federal Census provides agricultural statistics for both purposes every 5 years, whereas the BAE provides current, up-to-the-minute agricultural statistics. Actually in the field of agricultural statistics we have "3-way cooperation"—The BAE, the Census Bureau, and the States Agencies.

This cooperation could well be extended, not only in the field of current agricultural statistics, but also in providing the data needed for the agricultural economic and marketing research described by Dr. Elliott this morning. (The soundness of the conclusions reached in economic investigations depends in part on the reliability and the representativeness of the statistics and data utilized by the research worker.) Much of the data required for economic analysis relate to the operation and business activity of some 6,000,000 individual farms and/or to thousands of buyers and processors of agricultural products. Frequent enumerations of so large a population are not practical because of their high cost and lack of timeliness. Sampling in one form or another is the only alternative to complete enumeration. Since we cannot escape sampling, we should know more about it and how to use it effectively.

Sampling—A Universal Problem in Modern Science

The problem of sampling and of drawing inference from sample observations is a universal problem in modern science. All of the sciences face the problem of generalizing the results obtained from an analysis of selected observations, to the phenomenon as a whole, i.e., generalizing from a sample to an estimate of the whole. Fundamentally it is a problem in estimating or forecasting either in space or in time or in both.

In the field of the physical sciences where the phenomenon can be brought into the laboratory for more or less precise observation and experimentation, the basic problem of sampling is recognized and objective methods are being developed to minimize the need for intuitive judgment in drawing conclusions.

For example, quality control, a phase of modern statistical sampling, is now recognized as essential in the manufacture of precision articles produced by mass production methods, such as military supplies, ammunition, electric light bulbs, parts of telephone and electronic equipment. Large and well equipped laboratories are maintained for the purpose of quality control. The statistical technique of sequential analysis is a comparatively recent development in the field of quality control.

A beginning in the use of the sampling concept has been made in the field of meteorology which incidentally has much in common with the social sciences in that the phenomenon of interest cannot be brought into the laboratory. Also in a recent survey made by the American Society of Civil Engineers of its membership, concerning the phases of college training that need strengthening most if the graduates of our engineering colleges are to meet the demands of the job when they finished college, some 35% of the returns specified the need for more thorough training in modern statistical methods. Most of you are familiar with the rapid strides that have been made in recent years in developing statistical tests for determining the effectiveness of insecticides, fungicides and of other chemicals and drugs used for specific physiological purposes.

You are all aware, at least in a general way, of the development of efficient statistical designs for field plot and animal experiments - experimental designs that materially increase the amount of information obtained from a given expenditure of labor and funds. A statistically sound experimental design and method of sampling within plots makes it possible for the researcher to apply valid tests of significance to the results which materially reduce the necessity for depending upon uncertain intuitive judgment in drawing conclusions. He can, of course, choose the level of significance that he considers appropriate to the phenomenon being investigated. As a result he can state his conclusions and recommendations in more positive terms.

One of the basic trends in modern science from pure physics on the one hand to say psychology on the other is to increase the element of objectivity and decrease the area in which dependence is placed on intuitive judgment. Intellectual ingenuity and resourcefulness is utilized to the utmost by the investigator in developing reasonable hypotheses on the basis of available knowledge, and then these hypotheses are tested as rigorously as possible utilizing the most accurate and representative observational data obtainable.

In economic and other social science research we cannot bring the phenomenon that we wish to investigate into the laboratory for study, nor can we ordinarily conduct experiments in field plots. We can only take observations of the phenomenon in question as it changes in space and in time, and use these data in testing our hypotheses.

Ten Years of Progress in Developing Valid Methods of Sampling Economic Phenomena

During the last 10 years substantial progress has been made in the development of valid methods of sampling. In fact it was 10 years ago last month when Professor R. A. Fisher was teaching in the summer school of Iowa State College, that a week's conference of Corn Belt State Statisticians and Statistical Economists from the Land Grant Colleges was held in Ames. The problem for consideration at this conference was "How would one select a sample of farms that would be representative of all the farms in a State, in a type of farming area, or in the country as a whole." Many different opinions were expressed but none of them seemed very convincing. It was apparent that fundamental statistical research in sampling was required if a satisfactory solution was to be found. Consequently resolutions were adopted by the conference recommending that such research be initiated and carried on aggressively.

During the winter of 1936-37 a beginning was made by the BAE, financed in part by funds furnished by what is now known as the Division of Statistical Standards of the Bureau of the Budget. An informal interdepartmental committee was organized, made up of representatives of the Census Bureau, Bureau of Labor Statistics and BAE and Fredrick F. Stephan, Executive Secretary of the American Statistical Association, Morris Hansen was on this committee. J. J. Morgan was in charge of this first study in sampling. At the suggestion of Mr. Pottit of the Census Bureau, empirical tests were made, using Census "minor civil divisions" as sampling units. As was expected, the MCD did not prove to be an efficient size of sampling unit in estimating county, crop reporting district or state averages and totals, or percentage change from one census to the next. The MCD is too large a concentration of farms in the sampling unit and does not provide sufficient geographic distribution of the sample.

In 1937 a Bankhead-Jones research project in sampling was established in cooperation with the Statistical Laboratory of Iowa State College under the immediate direction of Professor Snedecor and Arnold J. King. In 1939 the College brought Professor William Cochran of Rothenstead Experimental Station to the Laboratory. (In the meantime a number of promising young graduate research assistants were taken on part time by the Laboratory—Earl Houseman, and Ray Jossen were members of this group.) One of the early and highly successful pilot projects in enumerative sampling was a two-year "matched sample" of about 400 farms selected to represent Iowa's approximately 200,000 farms. It gave a highly accurate estimate of the change in hog number. The preharvest wheat survey in the Great Plains States was another.

Many of you are familiar, at least in a general way, with the excellent work done by the Laboratory in cooperation with the Bureau of the Census and BAE in developing our present knowledge of the principles and practice of sampling. The Master Sample designed to provide a cross section of farms in this country was developed jointly by the Iowa State College Statistical Laboratory, the Bureau of the Census and BAE. Robert Straszheim of Indiana in the next paper will describe the Master Sample.

FEB 1 8 1947

The Bureau is continuing its cooperative research in statistical methodology with the statistical laboratories of the State Colleges of Iowa and North Carolina. The Bureau of the Census also is continuing its cooperation with the laboratory at Ames and with the BAE. Both laboratories have as their major function research in fundamental statistical methodology and underlying mathematical theory. The laboratory staffs deal with statistical methodology as generic. Some of the investigations deal directly with sampling problems of interest to the social sciences, and some relate primarily to problems encountered in biological work. Representatives of both laboratories will describe their fields of work on Friday morning of this week.

With 10 years of research and considerable experience in applying objective methods of sampling, not only in this country but also in parts of Europe, a body of knowledge has been acquired that gives us confidence that we can, with further experience, eventually solve the problem of selecting an efficient, representative sample of farms for any specified agricultural area, or for the United States as a whole. This is the problem that was seriously discussed at Iowa State College 10 years ago this summer.

Psychological Factors in Sampling

Although the statistical problem of selecting an efficient representative sample, one that will minimize error in the statistical sense, and at the same time minimize cost is probably our most important and fundamental problem, it is not, however, the only important problem confronting us at this time. All of you who have participated in taking a schedule or "farm record" by the interview method are aware of the mistakes and errors (non-sampling errors) that can creep into the record. These errors can be caused in many ways. Some are caused by the way the question is stated by the interviewer, by how the respondent understands the question; by the frame of mind or the mood of the respondent; or his inability to remember events or facts that occurred sometime in the past; if the schedule is too long, by fatigue on the part of both the respondent and interviewer and vague instructions and definitions of concepts. Lack of training and familiarity with the subject on the part of the interviewer or his supervisor all contribute to these "human errors" as contrasted with sampling errors.

In seeking a solution of this problem of minimizing "human errors", we need to draw on the basic principles of psychology, sociology and related social sciences that deal with an understanding of human behavior and human weaknesses. The Division of Special Surveys of BAE has made phenomenal progress during the last 6 years in meeting many of these psychological problems that arise in connection with enumerative and interview sampling. The two leading psychologists of this Division, Rensis Likert and Charles Cannell will discuss some of these problems Thursday. They are problems for which we must draw upon our own experience and so called "common sense" in seeking solutions. They also are problems on which a great deal of research and experimentation is needed to determine basic principles and efficient techniques. I think that we all recognize that a chain can be no

stronger than its weakest link. Respondent and interview bias are weak links in this chain today. Research to strengthen these weak links will be undertaken soon.

Choice of Sampling Methods

What choice of sampling methods does the agricultural statistician and the research economist have? What considerations should be weighed and balanced in deciding on the method of sampling to be used with a specific problem?

We are looking forward to the time when both the agricultural statistician and the investigator of agricultural economic and marketing problems will have sufficient funds to be in position to utilize whatever method of securing his basic observational data as will provide the necessary degree of accuracy at a minimum cost. If the number of individuals in the population that forms his "universe of inquiry" is small, and these individuals show a high degree of variability, either in space or time, and there is need for great accuracy, a complete enumeration is undoubtedly the most satisfactory method for obtaining "the required data. Such might be the case if he wanted to determine the distribution of fruit trees by kind, variety, and age in an important commercial orchard area. Another example might be a marketing study where the total number of processors or handlers is relatively small and their plants vary greatly as to capacity.

At the other extreme, the investigation might require a small "hand picked" sample or judgment sample of farms or processors for detailed case history investigations. It can be readily demonstrated that with a very small sample, probably less than 5, a judgment sample based on rather complete but general information will provide a more representative sample than stratified random selection techniques. However, as the size of sample increases, it is surprising how soon random selection provides a more representative and dependable sample.

It is in between these two limits of the complete enumeration on the one hand and a hand picked sample of a very few observations for case history studies on the other, that modern scientific sampling becomes a necessary and valuable aid to both the statistician and the economic investigator. It is here that the choice of the method of sampling becomes fully as important to him as is the choice of the experimental design to the agronomist. In fact the design of an experiment and the design of a sample both stem from a consideration of the same fundamental concepts in mathematical statistics. Broadly speaking he has a choice of the three primary alternatives of (1) a mailed sample, (2) a preselected interview or enumerative sample, or (3) a combination of mailed and interview sample.

1. Mailed Sample: Generally speaking if timeliness is of paramount importance, absolute accuracy secondary, and reliable check data such as market receipts and "bench marks" from census data are available historically, one can frequently rely on a mailed sample provided no sharp or far reaching changes are taking place in the underlying physical and psychological situation that may cause the bias or representativeness of the mailed sample to change in time. One can seldom tell in advance when and under what conditions these biases will change, or in what direction, or how much.

Most of the sampling done in the preparation of crop and livestock estimates has been by mailed inquiries to lists of correspondents maintained by the State Statisticians or through the Rural Mail Carriers. Fortunately by utilizing the census data as a bench mark every 5 years, and market receipts, cotton ginnings and other check data, including Assessor's annual farm censuses of crop acreages in about 10 or 11 important agricultural States, and livestock assessment data in more than half of the States, reasonably satisfactory State and National estimates have been made for many important items. The comparison of crop estimates with census data for 1944 substantiates this conclusion.

2. A Presselected Interview or Enumerative Sample: In the broad field of agricultural statistics and research there are many phenomena that experience has shown cannot be sampled satisfactorily by any form of a mailed inquiry. Either the type of schedule is necessarily so long and complex that an interviewer is required to explain what is wanted to the respondent and to "keep at" him until all the information is obtained, or the subject matter of the inquiry is such that correspondents who will reply voluntarily by mail are highly selective. It is with phenomena of this nature that the presselected interview sample should be chosen.

With this type of sample, valid and efficient sampling methods are used in designing the sample. The farms or other units of observation are presselected for the interviewer making the survey. The interviewer exercises no direction in choosing the respondents that comprise the sample. The farm on the muddy road, the farm with the rundown buildings as well as the prosperous looking farmstead, all have a known probability of being included in the sample. By the same token, friends of the interviewer-- those belonging to his church, his farm organization or his lodge-- are no more likely to appear in the survey than persons belonging to any other group.

The major advantage of the respondent preselection procedure is that it is mathematically and statistically sound. For this reason it taps a vast source of knowledge, both in the science of statistics and mathematics and in the fields of science where statistical methods have reached a high state of practical application. A transfer of these techniques from other fields makes it possible to provide surveys of increasing accuracy at decreasing cost.

The use of basic sampling methods which are statistically valid is a matter of deep significance to agricultural economic and marketing research. Statistically valid sampling methods combined with properly trained and supervised interviewers makes it possible to determine the degree of accuracy of the survey results. This can be done with no other method of sampling.

In the case of the proposed January 1947 Enumerative Sample of 15,000 farms selected to provide a representative cross section of American farms, the method of selecting the sample is as follows: (It will be described more fully at the Thursday afternoon meeting.)

- a. The agricultural area of the United States is divided into about 400 homogeneous areas or strata consisting usually of several counties each. The stratification is based primarily on the type of farming areas modified to some extent in order to provide approximately the same number of census farms in each. These strata with only a few exceptions are all within the boundaries of a State.
- b. Within each of these 400 strata one county is selected at random, and
- c. A part-time enumerator will be hired in each of these counties or 400 in all. In important agricultural strata, one additional county usually will be selected at random for inclusion in the sample, and in a few cases two additional counties, thus making a total of perhaps 900 sample counties.
- d. Within each of these 900 counties, segments of the Master Sample will be selected at random and the interviewer will be expected to enumerate the farmers having farmsteads in these segments. Detailed maps and aerial photographs will be furnished the interviewer to enable him to establish the boundaries of these sample segments. There are from 4 to 6 farms in each sample segment.
- e. With a 15,000 farm sample, from 15 to 20 farms will be enumerated in each of these 900 counties.
- f. The 41 State Agricultural Statisticians' offices will select, train and supervise these enumerators. The person in each of these offices who will supervise this survey will receive a week of instruction at one of three area conferences prior to December 15, 1946.

This brief description of the January 1947 Enumerative Survey shows specifically what is meant by a Preselected Interview Sample. It will provide a representative, general purpose sample of farms for each State, as each farm in the State has a known probability of being included in the sample. Although representative, it is not necessarily adequate as to size on a State basis for most purposes. A substantial increase in size would improve its adequacy and reduce the sampling error.

3. The remaining alternative method of sampling involves a combination of mailed sample and preselected interview sample. This combination method would be appropriate where the mailed sample has measurable validity but cannot be depended upon to produce results of needed accuracy. The two types of samples could be taken simultaneously with the interview sample in say 20 to 50 percent of the geographic areas and the mailed sample in all areas and regression (correlation) methods used to correct the mailed sample for bias. The mailed sample would have a more uniform scatter over a State or region, whereas the interview sample would be taken in selected counties. The non-respondents from the mailed sample may be sampled by the interview method. These and other ways of combining these two methods of sampling will be brought out by Walter Hendricks and Arnold King on Thursday.

It might be well at this point to bring out more clearly the considerations that must be carefully weighed and balanced each with the other by the statistician or research economist in arriving at a choice of sampling method to be used in connection with any specific problem. The more important considerations or factors are as follows:

1. The first and frequently the most important consideration is the funds and resources of manpower available for the job. Money and time are usually the most critical limiting factors in sampling. The desire for accuracy and a sound basis for generalization is usually, but not always, practically unlimited.
2. The degree of accuracy required in the final results whether these results are estimates, or conclusions obtained from economic investigations and research.
3. The nature of the quantity to be estimated, whether the absolute level or amount on the one hand or relative accuracy such as the change of level from one period to the next, expressed as a percentage on the other. For example the number of hogs or bushels of corn produced, the year-to-year change in the number of hogs or production of corn, or within season variation of prices or marketings.
4. The time available. Often timeliness is more important than a high degree of accuracy. This is especially true with estimates and forecasts of short season commercial fruit and vegetable crops. A rough approximation of the change in supply due to frost damage on perishable crops, promptly made, usually has greater economic significance than a much more accurate estimate that comes out several weeks or months later.
5. Availability of reliable check or "bench mark" statistics such as total ginnings of cotton and marketings, truck and car-lot shipments of fruit and vegetable crops, slaughter of livestock, etc. If reliable check data are available, the statistician can get along pretty well with a sample that shows relative change accurately. This is one of the reasons why the mailed samples have proved so useful for items for which dependable check data are available.

6. Inate bias in the observations themselves. In the case of cash crops before they have left the producer's hands there is a general tendency to obtain reports, whether by mail or interview, that are understatement of the true acreage or production. In such cases a sample of direct plant observations is the only sound approach, such as boll counts of cotton, head samples of wheat, ear samples of corn, or the use of the crop meter to measure change in crop acreages from year to year.
7. Extent to which generalization from the sample data is thought desirable and necessary. While the investigator may think he can limit the application of his conclusions to the observed cases, perhaps farms that he has selected for his survey, he cannot prevent others from generalizing and using these results far beyond what he originally intended.
8. Fund of knowledge already available to the investigator. If very little is known about a specific or perhaps complex phenomenon, it may be preferable to make a detailed case history study of a few hand-picked cases rather than obtain a representative sample; or the investigator may be satisfied with most any kind of a sample in order to get some data with which to work and to serve as a basis for more scientific investigations later.

In the field of Agricultural Estimates the first 5 considerations have been of paramount importance. Funds and resources have never been adequate to obtain estimates with the degree of accuracy and timeliness desired by the statistician, the economist and others using his data. Consequently the agricultural statistician has been compelled to use mailed samples with a varying amount of subjective interpretation on his part. The mailed sample is usually the least expensive method of sampling and the schedules can be mailed out, returned and summarized with sufficient dispatch to meet rigid dead lines on releases. Fortunately, with census data "bench marks" and other check data from commercial sources, he has been able to utilize a method of sampling which usually showed change more accurately than absolute values.

With crops that producers tend to habitually understate prior to harvest, the agricultural statistician has shown considerable ingenuity in developing statistical methods for correcting the bias in the sample using census, assessor, or commercial marketings; he has learned how to get information from key persons familiar with the crop locally and how to evaluate it for bias. He has developed methods for observing the growing crop in the field by use of the crop meter and plant counts and measurements.

Experience with Enumerative Surveys.

The Division of Rural Population in cooperation with Agricultural Estimates has conducted four nation-wide enumerative Farm Wage and Labor Surveys during the last year and a half-- March, May and September 1945 and July 1946-- and also special Farm Wage and Labor Surveys in 60 commercial crop areas covering 74 crops in 15 different States. The latter were designed to obtain wage rates received by the large number of seasonal workers employed in harvesting certain commercial crops and also other specific information such as age and sex of workers, hours worked, source of workers and piece rates.

The sample for the four nation-wide Farm Wage and Labor Surveys was designed to provide for national estimates and estimates by the four major regions of the United States. It was not designed as a basis for estimates by States or small groups of States. The sample of about 20,000 farms was taken in 158 sample counties; one or more sample counties were selected in each of 101 strata. These strata were determined on the basis of the 10 major type of farming areas and the four major regions of the United States.

The Division of Special Surveys, under the able leadership of Ronsis Likert, has had broad experience since 1939 with interview surveys focused primarily on attitudes, but more recently in cooperation with Agricultural Estimates with enumerative surveys. The objectives of the latter have been to obtain economic data from a sample of preselected farmer respondents.

The Quarterly Surveys of Agriculture (QSA) made in April, July and October, 1945 and January 1946 are probably of greatest interest to the persons attending this conference. The primary objectives of QSA may be summarized as follows:

1. To see what could be done by enumerative sampling in getting information on topics and items that cannot be obtained satisfactorily by mailed samples or any other means.
2. To obtain a sample from which national estimates could be made for comparison with official BAE estimates.
3. To obtain data on individual farm operations for use in economic analysis.
4. To gain experience in conducting a nation-wide enumerative survey of economic topics.

A sample of 101 counties, one in each of 101 strata mentioned above, was used for the QSA with about six Master Sample segments within each of the 101 sample counties. A total sample of nearly 3,000 farms was planned. The same farms were visited each of the four quarters except as farms dropped out of the survey. The number of useable farm schedules obtained was as follows:

| | | | |
|------|---------|-----------|-------|
| 1945 | April | - - - - - | -2803 |
| | July | - - - - - | -2625 |
| | October | - - - - - | -2530 |
| 1946 | January | - - - - - | -2345 |

The results obtained from the QSA were about what the statisticians expected from a small nation-wide sample. Less than 0.05 percent of the total number of farms in the United States were included in the sample. Its geographic distribution was limited to 101 counties, or about 3 percent of the total counties in the United States; in fact the over-all sampling rate was slightly less than one farm per agricultural county. Furthermore the variability of farms over the entire United States is high.

With such a sample one could expect to obtain reasonably valid national estimates and frequency distributions only on phenomena that occur on a high proportion of farms or with aggregates of several less frequently occurring items. A breakdown by the four major regions would not be practical. However, the preliminary results are interesting.

On rental arrangements taken in April 1945 the QSA showed 32.5 percent of the farmers as full tenants of share croppers and the 1945 census 32.3 percent.

Total acres in crops in the QSA was slightly higher than was shown by the census.

Land in farms in QSA was somewhat lower than the census.

Cash farm income was taken quarterly along with inventories of livestock and crops. A national estimate of livestock and poultry sales was about 20 percent below the BAE estimates; livestock and poultry products was about 15-20 percent lower; and crop sales were about 6 to 8 percent low as compared with BAE estimates. The aggregate of these three sales was about 10 percent less than BAE estimates. In the case of non-farm income the QSA estimates were about 10 percent below the BAE estimates.

Although no complete check has been made of cash farm expenditures, our impression is that the estimates from the QSA sample may be in excess of the true values. Information on market outlets for eggs, live chickens and hogs was obtained each quarter. The data appear to be reasonable and useable when the four quarters are added together. It would appear feasible to compile these data in three categories-- sold direct to consumers, to local markets, and elsewhere.

There is little evidence that stated intentions to market cattle are reliable indications of future marketing on a nation-wide basis. Most of the farmers who indicated that they would sell cattle did sell but about an equal number of farmers, who said they would not sell, did sell. The reasons for this behavior could not be obtained satisfactorily by part-time interviewers. Intentions to sell appear to be least dependable with milk cows and most dependable with fattening steers and heifers. The sample was too small to permit special tabulation for the Corn Belt States where intentions to sell cattle might have more meaning.

Comparisons of household facilities with 1945 census indicate a very close correspondence.

The information obtained on insurance carried by farmers does not appear to be reliable. The questionnaire construction was poor and it is very difficult to obtain accurate information on this topic because of the many forms in which insurance is written.

A check on short-term debts indicates that the QSA was about 50 percent short of official BAE estimates, however the relative amounts of short-term debts owed to different types of lenders were probably quite accurate. The data on farm mortgage debts have not yet been tabulated.

Most of the estimates on financial assets and investments are about one half as large as regular BAE estimates, however, regular BAE estimates cover farm population while the QSA covers only farm operators. Again the relative proportion of assets appears reasonably accurate.

The data on medical care received by farm families were new information and check data are not available, however, the results appear to be reasonable.

Administratively the experience with the QSA was not satisfactory. The schedules contained so many different topics and such a large number of questions to be answered by the respondent, i.e., "respondent items," that it was not possible to adequately train the supervisors and interviewers. The length, complexity and nature of the schedules were such that many interviewers and some supervisors quit the job and it was extremely difficult to replace and train new ones. As a result, overhead and interview expense was excessive.

Although actual interview time ranged from 45 minutes on small farms to an hour and a half or two hours on large farms, the enumerators averaged only about 1.8 schedules per 8 hour day. Much time was consumed in locating respondents and in making "call backs" when farmers were not at home the first time the interviewer called. For example, the cost of part-time county supervisors and enumerators for salaries and travel averaged nearly \$5.00 per useable schedule for the April QSA.

The coding, editing and punching of the cards for the QSA was also a slow, tedious job that has only recently been completed for the January 1946 QSA. Forty-six punched cards are required for the data on each farm.

Our experience with the QSA indicates conclusively that the economists of a research organization cannot get all of the information they would like to have from one survey or with one schedule even when the survey is spread over the four quarters of the year. The greater the number of different "topics," or subject matter fields included in a survey, the more difficult is the matter of training and supervising of non-experienced part-time interviewers. Each new topic must be introduced by the interviewer for the purpose of establishing the necessary rapport with the respondent. Consequently the greater the number of topics, the longer the interview time and the higher the cost per schedule. Furthermore, topics that are likely to cause resistance on the part of the respondent should be avoided, at least until the interviewers become more experienced and specially prepared instructions and training can be provided. Dr. Likert and his staff have demonstrated that these difficult topics can be handled by interview methods by well-trained and experienced enumerators.

Enumerative Sampling Program

Objectives: The primary objectives of the Bureau's program for enumerative sampling are:

1. To make full and effective use of the constantly improving techniques of sampling in meeting national and State needs for more reliable and useful agricultural statistics and for representative economic data for analysis and research. The sampling techniques to be used in each instance are

those that will provide maximum accuracy at minimum cost. From the research in sampling now underway and from the experience to be gained in this sampling program, it is expected that more powerful techniques will be developed that will combine the advantages of the mailed sample and the preselected enumerative sample.

2. To encourage and develop two-way cooperation between the Bureau and State agencies-- Land Grant Colleges and State Departments of Agriculture-- in the entire field of agricultural sampling for the purpose of placing a more sure foundation of methodology under current agricultural statistics and agricultural economic research. Such cooperation will be mutually advantageous. The Bureau is primarily interested in a representative sample that is sufficiently adequate as to size to provide reliable State statistics and also dependable estimates for areas within the State such as type-of-farming areas and even counties. We all need the same basic data concerning agriculture and marketing. A pooling of needs, ideas, technical "know how" and financial and personnel resources can be made to pay handsome dividends to American agriculture and to the taxpayer, who after all is our employer.

Organization: The Bureau, with its 41 State Offices, is in position to provide the technical direction and supervision necessary on a nation-wide basis to insure comparability in estimates based on sample data. The agricultural statistician in each of the Bureau's State Offices are responsible for the actual collection of sample data, whether it be a mailed or enumerative sample.

As soon as the sample counties in each State are selected, the State Offices will select, hire and train the local part-time interviewers, and directly supervise their work of enumeration of farms in the selected Master Sample segments. The number of strata and the number of part-time interviewers per State varies from one in Arizona and Wyoming, 11 each in Michigan, Indiana, Virginia, Alabama and Georgia, to 22 in Texas.

Later the State Offices will supervise the editing and coding of the schedules before they are sent to the Washington Office for entering on punch cards, machine tabulation and summarization. Once the data are on punched cards, a duplicate deck of cards can be furnished cooperating State agencies having IBM equipment.

In those States which desire a larger sample for the purpose of within-State analysis or estimates for areas within the State, the sample can be made as large as State sponsored funds and other needed assistance will permit. It merely involves an expansion of existing machinery.

Sample Design: The Bureau is now designing a nation-wide, general-purpose sample that will be suitable for use in making estimates by States for items commonly occurring on farms. The new sample is highly flexible as it can be contracted or expanded, or changed from year to year with minimum expense and difficulty. As already explained, the new sample for use with the January 1947 enumerative survey will consist of 4 or 5 Master Sample segments, of 15 to 20 farms each, in each of nearly 900 counties, or a total sample of about 15,000 farms for the United States. The QSA sample of less than 3,000 farms was taken in 101 counties.

In selecting the sample counties, all the agricultural counties of the United States were grouped into 400 homogeneous areas or strata with about the same number of farms in each using the type-of-farming area classification. In only a very few instances do these strata include counties in two adjacent States. One county was selected at random in each of the 400 strata. The 400 part-time interviewers will be selected from among persons living in these 400 counties. Each interviewer will enumerate the Master Sample segments in his home county and call-backs will be limited largely to these counties. In addition, nearly 500 other counties will be selected at random in these 400 strata, one or two in each strata. The part-time interviewer also will enumerate the Master Sample segments in these additional 1 or 2 counties in his strata.

The importance of getting a wide distribution of the sample among counties is appreciated when it is realized that a 15,000 farm sample in 800 counties will be about equal in statistical reliability to a 20,000 farm sample taken in 400 counties.

The flexibility of the new sample design is readily apparent. It could be reduced to the 400 counties in which the interviewer has his home, or it could be expanded to 1300 or 1700 counties by the addition of one or two sample counties in each of the 400 strata. The size of the sample also can be expanded by increasing the number of Master Sample segments to be enumerated within the sample counties.

Frequency of Sampling: The Bureau is planning to take enumerative surveys quarterly—early in January, April, July and October. Because of the limited funds now available for this work, the January 1947 survey of 15,000 farms will be a major survey and the one in April a minor survey with probably only about 7,500 farms. The schedule also will be shorter in April in order to keep the expense to the lowest possible amount. It is hoped that funds will be available to permit a 15,000 sample in July and a much larger major sample in October, 1947. On the basis of experience to date and need for the information, there would appear to be ample justification for at least one and probably two major surveys and two minor enumerative surveys per year. The timing of the major surveys might vary geographically because of the seasonal nature of farm operations such as planting, harvesting and marketing.

Some types of agricultural and marketing phenomena change so slowly in time that data are needed only biennially or at longer intervals or perhaps the census can get the data every 5 years. Other types are so dominated by year-to-year influences that they are needed annually. Still others are so closely associated with the "march of the seasons" that observational data are required quarterly or at least very soon after the events occur in order to minimize memory bias on the part of the respondent. Timeliness is also important from the standpoint of the usefulness of the information.

Many of our important economic statistical series are on an annual basis and January coming as it does immediately after the close of the year is the best time to get certain information for a full calendar year such as cash farm income, non-farm family income, farm expenses, and farm population. January

is also the accepted month in many areas for taking inventories of livestock and machinery. On the other hand, information concerning land use and acreages of various crops, can probably be obtained best in October. There are other items with high seasonal variation that should be gotten quarterly, such as farm labor and wages, milk and egg production and stocks of grain on farms. There are other classes of items that can be enumerated satisfactorily in any of the 4 quarters of the year.

In view of the many demands for enumerative survey data and the physical fact that each survey schedule must be kept within practical operating limits, it will be necessary to work out an integrated annual and quarterly sampling scheme extending over several years, one that will utilize both enumerative and mailed sampling techniques singly and in combination.

Selection of Topics and Items: The subject matter fields or "topics" for which representative samples are needed by the agricultural statistician and research economist are wide and diverse. The major lines of data which should be collected primarily through enumerative survey procedures are listed on pages 32 and 33 of the Proposed Program of Research and Statistics of BAE discussed by Dr. Elliott this morning. This list, with some modifications, is as follows:

1. Agricultural employment and wages paid to hired agricultural workers, including perquisites. -- Since funds have been directly appropriated for this topic, it will be included quarterly.
2. Farm accidents and injuries causing loss of time. -- Very little reliable information is available anywhere on this important topic which so vitally affects all farm workers. For at least a year this topic will be included quarterly.
3. Farm population and migration. -- This topic is pretty well covered from other sources, but farm population is needed as a control item for analysis, especially with topic 2 above.
4. Acreage and production of crops, stocks, livestock inventories and production, production of milk and milk products, poultry and eggs, wool and honey, slaughter of livestock, prices paid and received by farmers, land values. -- These are all basic agricultural statistics, some of which could undoubtedly be strengthened by enumerative sampling or by a combination of mailed and enumerative sampling.
5. Farm operating expenditures by size distributions. -- This topic is needed currently along with farm wages to obtain a reliable picture of costs for American agriculture. It should be included annually and perhaps quarterly.
6. Cash income of farmers and farm families, levels and size distributions. -- This information is basic to agricultural economics and is also needed to complete the record for national income. It should be included annually and perhaps quarterly.

7. Expenditures and savings of farm families. -- This is a time consuming and difficult topic to handle by any method of interview sampling. Two respondents are involved, the farm operator and his wife.
8. Income distributions of farm laborers' families.
9. Number of all farms, and number of farms reporting major crop and live-stock items. -- The use of area sampling, i.e., Master Sample segments should enable us to obtain this information.
10. Tenure and land ownership.
11. Gross and net rent payable on farm lands and landlord's share of government payments.
12. Farm practices, including availability and use of modern technological devices and procedures.
13. Amount and characteristics of the various types of credit used by farmers. -- This topic requires well-trained and experienced enumerators.
14. Methods of marketing.
15. Goods and services secured without cash expenditures.
16. Availability and use of goods and services not adequately measured in terms of cash expenditures.

The above list is long and includes many topics which require considerable elaboration. Schedule construction is difficult with several of them, especially a schedule that will minimize interview time and friction with the respondent and at the same time obtain the information required. A number of these topics should be tried as a "pilot project" in a limited area before being placed on a State or national survey. Not all of the topics on the list are required in the same degree of accuracy, with the same frequency, for the same geographic areas, or in the same detail.

Enumerative Surveys for 1947:

The topics planned for the January 1947 schedule include:

1. General
2. Farm Tenure and Rent.
3. Value of Farm
4. Farm Dwellings and Population
5. Farm Wages and Employment for the week ending January 11, 1947 and expected number of man-days of hired labor that will be required during 1947.

6. Accidents involving loss of time during October, November & December 1946.
7. Livestock Numbers, January 1, 1947 -- cattle, cows, cattle on grain feed and marketing intentions, and workstock.
8. Farm machinery inventory-- age and size or model of automobiles, tractors, combines, corn pickers, multiple row cultivators, windrow pick-up balers, milking machines and value of farm machinery.
9. Cash Farm Expenses, Operator and Landlord.
10. Cash Farm Receipts, Operator and Landlord.
11. Other Cash Income, Operator
12. Family Living Expenses -- Total only to be obtained from the operator.

This schedule will contain about 25 percent fewer "respondent items" than the average QSA.

The topics planned for the smaller April 1947 Survey of about 7,500 farms on a subsample of the January Survey are about as follows:

1. General
2. Farm Population
3. Farm Wages and Employment for the week ending April 12, 1947.
4. Accidents involving loss of time during January, February and March 1947.
5. Milk and egg production yesterday.
6. Cattle on grain feed and marketing intentions.
7. Farm Grain and Feed Stocks, April 1, 1947
8. Winter wheat acreage planted and remaining for harvest

Some of the following topics also may be included:

9. Housing and building construction
10. Medical care and availability of facilities.
11. Education expenses.
12. Farm practices and use of modern technological devices and procedures.
13. Agricultural prices.

In July the first six topics planned for April will be repeated and some of the following topics included:

1. Livestock items carried on June Rural Carrier Livestock Survey.
2. Farm Mortgage debt, interest rate and type of lender.
3. Short-term credit.
4. Farm practices including use of fertilizer and lime, conservation practices, regionalized to some extent.
5. Methods of Marketing.
6. Agricultural prices.

In October 1947 it is planned to take a major sample for which the schedule would, in addition to the first six topics used in April and July, place primary emphasis on such topics as --

1. Land utilization
2. Acreage and production of important crops
3. Livestock items designed to get size of fall pig crop and breeding intentions.
4. Land ownership
5. Farm field practices and use of modern technological devices and procedures.
6. Agricultural prices.

On Thursday afternoon further consideration will be given to the January 1947 Enumerative Survey -- design of the sample, selection and training of interviewers, and minimizing interviewer and respondent bias and refusals.